DOCUMENT RESUME

ED 279 689 TM 870 076

AUTHOR Newmann, Fred M.
TITLE The Assessment of Discourse in Social Studies.
PUB DATE 15 Aug 86
NOTE 22p.; One of 46 papers commissioned by the Study Group on the National Assessment of Student Achievement and cited in Appendix B to their final report "The Nation's Report Card" (TM 870 049). For other papers in this group, see TM 870 050-094.
PUB TYPE Viewpoints (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Discourse Analysis; *Educational Assessment; Educational Testing; Elementary Secondary Education; *Essay Tests; Measurement Objectives; National Surveys; *Social Studies; *Speech Communication; Test Format; Testing Problems; Verbal Tests; *Writing Evaluation
IDENTIFIERS *National Assessment of Educational Progress; Thinking Skills

ABSTRACT

Social studies assessment for the National Assessment of Educational Progress (NAEP) should focus on students' oral and written discourse on social topics. Discourse is language produced by the student with the intention of providing a narrative, argument, explanation, or analysis. The assessment of discourse is important because: emphasis on discourse will facilitate social studies learning, this type of assessment indicates higher level thinking, and the development discourse itself is the ultimate objective in social studies education. Knowledge need not be tested first, separate from the discourse intended to clarify a problem or topic. The production of language is an integral part of learning and does not occur only after bits of knowledge are acquired. Instead, the acquisition, storage, retrieval, application, and synthesis of knowledge are all complex and interactive processes. Topics which would be suitable for testing students' ability to conduct intelligent discourse, both written and oral, include: the Constitution and Bill of Rights, slavery, and citizenship responsibilities. A reasonable response would be a 300-word essay or an oral response which could be audiotaped and transcribed. Logistical testing problems and students' lack of classroom speaking experience both present difficulties for oral assessment. Holistic and primary trait scoring predominate as forms of direct assessment of students' writing skills; both types of scoring produce reliable scores at acceptable levels of cost. The challenge in discourse analysis lies in the area of comparisons across age groups and nations. Seven sample test items are appended. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED279689

The Assessment of Discourse in Social Studies

Fred M. Newmann

University of Wisconsin

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☑ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

F. M. Newmann

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper commissioned by

THE STUDY GROUP ON THE NATIONAL ASSESSMENT OF STUDENT ACHIEVEMENT

1986

TM 870 076

BEST COPY AVAILABLE

The Assessment of Discourse in Social Studies

Fred M. Newmann
University of Wisconsin, Madison

with the assistance of
Mary Louise Gomez
University of Wisconsin, Madison

August 15, 1986

DRAFT. NOT FOR DISTRIBUTION.

Paper prepared for the Study Group on National Assessment, US Department of Education.

This paper was prepared at the National Center on Effective Secondary Schools, School of Education, University of Wisconsin-Madison which is supported in part by a grant from the Office of Educational Research and Improvement (Grant No. OERI-G-86-0007). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author and do not necessarily reflect the views of this agency or the US Department of Education.

Introduction

This is a position statement arguing that a major aspect of national social studies assessment should focus on the oral and written discourse produced by students on social topics. To develop the position, several issues will be addressed.

1. What is discourse?

2. Why is discourse an important national indicator of student achievement in history and social studies?

3. What forms of discourse should be assessed and why?

4. What does experience and research suggest regarding the feasibility of this approach to assessment?

5. Will discourse assessment provide meaningful comparative indicators?

The recommendation is not that all traditional indicators be replaced by indicators of discourse, but that discourse should occupy a significant place in a scheme of indicators.

I What is discourse?

We define discourse as language produced by the student with the intention of providing a narrative, argument, explanation, or analysis. Discourse requires the use of many words and at least several sentences which the student tries to organize into a coherent pattern of speech or writing. In producing such statements, the student assumes that the success of the task rests not simply on including a set of isolated words, numbers, or phrases within the statement, but creatively pulling knowledge together in the students' own language; that is, integrating into the students' language relevant formal or authoritative knowledge to which the student has been exposed (either through schooling or other experiences). Here are some hypothetical examples of student discourse.[1]

---

[1]Most discourse analysts (e.g. Halliday and Hasan, 1976) define discourse as any natural language occurrence or utterance in which a speaker/writer tries to communicate with a listener/reader. According to such a broad definition, the number of words or other linguistic criteria are not significant (e.g. an EXIT sign or a short question would count as discourse). To emphasize the importance and power of language that elaborates human experience, we have added criteria to the formal definition.

a. The Boston Tea Party before the American Revolution was a violent protest against the many restrictions on the colonists' trade imposed by the English Parliament. Although it was "violent," it really didn't hurt people - physically. It did do violence to property rights, because all that tea was destroyed by people who didn't own it. Still, I think the protest was justified, because the colonists had taken just about all the peaceful means they could to try to get Parliament to revise some very unjust laws.

b. We got into the Great Depression, because people lost faith in the economy. When people lost faith that prices of stock would rise, they rushed to sell. This lowered prices, people panicked, prices fell further, more panic, etc. As companies lost assets, they laid off workers. The workers as consumers couldn't afford to buy, so companies had to lower prices. This reduced their income further which led to further layoffs, failure to pay bank loans, etc. As people needed their savings, they rushed to the banks. Banks couldn't pay and they failed, which further undercut investment in business. The fear of economic failure made it fail.

Statements like this stand in stark contrast to the forms of expression usually required in conventional assessment tasks. Most tasks on standardized tests and in NAEP exercises do not require students to make statements using language. Instead, they typically require students to read a question that can be answered only by the student's choosing the "best" answer from among about four alternatives previously constructed by the testmaker. Such questions fail to generate discourse, because they offer no opportunity for the student to produce language.

II Why is discourse an important national indicator of student achievement in history and social studies?

The testing or assessment of student achievement can serve many different (and even contradictory) purposes. It may offer information about specific individuals (students, teachers) and/or information about groups and institutional aggregations (classes, schools, districts, states, nations). Individual or group level data can be used in diagnostic and formative ways to suggest specific strategies for improvement; it may be used to monitor progress toward goals; and it may be used for purposes of selection or the distribution of rewards and opportunities. We will not discuss here the various general purposes that NAEP assessments could or ought to serve. Instead, we assume that a major purpose is

to provide aggregate indicators (at the national, state, and possibly district or school level) of achievement in social studies. This purpose assumes that goals can be defined for the groups to be assessed, and that the main function of the assessment is to determine the extent to which goals have been achieved. Admittedly, if the assessment is constructed with only this purpose in mind, it is unlikely to serve other purposes such as providing feedback on what groups or individuals might do to improve performance.

The case for the assessment of discourse rests on three main points: emphasis upon discourse will facilitate the learning of content and the development of attitudes critical to history and the social studies; the assessment of discourse is a critical indicator of another important objective - higher level thinking; and most importantly, the ultimate educational objective in social studies (as well as in other subjects) is the development of discourse itself.

The effort to produce language about what one knows and understands should be seen not simply as a demonstration or a performance, but also as catalyst or process that enhances knowledge itself. Even in the taking of notes intended only for communication with self, as in a learning log, the creation of discourse enlarges understanding (Nystrand and Wiederspiel, 1977). This is especially apparent when we ask ourselves questions such as "What was the central point of the politician's speech? "Do I really understand how interest rates affect employment?" "How would I worship if I were a Moslem?" "Why did Lincoln try so hard to save the Union?" Producing coherent language to answer such questions helps to teach the content of social studies. To the extent that discourse is oriented toward communication with others, it also supports the development of students' abilities to take the role of the other (Nystrand, in press). To think carefully about how to make one's speech or writing communicate with others is to anticipate the concerns and perspectives of others. Discourse thus requires reciprocal thinking and attitudes which lay at the root of tolerance.

Assessment of discourse also provides unique opportunities to examine the quality of students' higher level thinking (Applebee, 1984). The promotion of higher order thinking has often been endorsed as a key objective within the social studies community, primarily for its connection to democratic citizenship. Critical thinking and problem-solving with regard to social concerns has been considered crucial to the preservation of democratic values such as consent of the governed, individual liberty and the public good. Higher order thinking has recently attracted increasing attention from psychologists, educational leaders and policymakers (Resnick, 1986). It has been tied not only to citizenship, but also to economic productivity on the grounds that problem solving skills are increasingly needed by workers as career changes demand new forms of

adaptation and as jobs become more complex, challenging workers not merely to find answers to problems, but also to identify the problems themselves and to create processes for developing solutions.

To date, however, there has been relatively little scientific attention devoted to assessment of higher level thinking. It has been tested through traditional formats, but a strong argument can be made that these have been inadequate (McPeck, 1981; Fredericksen, 1984, 1986), primarily because they give little information about how students integrate knowledge to solve problems. Asking students to define a problem and to produce a chain of reasoning or thinking in language gives richer information about the workings of their minds than asking them only to select choices from multiple choice-items. Traditional testing formats may tap some aspects of higher order thought, and not all forms of discourse can be considered indicators of it, but discourse indicators are necessary to expand our knowledge of student progress on this objective, especially with regard to thinking about ill-structured problems.

Authorities within the fields of history and social studies may differ on the particular forms of knowledge that students should learn and the kinds of problems to which they should apply their minds. In this sense, choosing proper indicators for achievement in social studies will always be problematic. Nevertheless, most rationales for history or social studies curriculum in US public schools are grounded in a conception of citizenship within a democracy. The most well-developed rationales (e.g. Oliver and Shaver, 1966) make a persuasive case that the ultimate indicators of the quality of citizenship are the statements that people make, whether orally or in writing, indicating their understanding of society, institutions, history, the relationship of citizens to one another and to the state. In short, we should look to discourse as an indicator, because our ultimate educational objective is discourse itself.

If time permitted more attention to the assessment of students' oral and written discourse in school classrooms, and if teachers felt they would not be challenged to prove the "objectivity" of these judgments, they say they would rely far more heavily on discourse as the most valid indicator of students' understanding. The power of discourse as an indicator of understanding is further illustrated by the emphasis placed in higher education on the writing of papers and essays and the use of oral examinations. Recent analysts of high school reform from Adler (1982), to Boyer (1983), to Sizer (1984) emphasize discourse as the ultimate indicator of an educated person. This seems so apparent, that it leads one to wonder why it has <u>not</u> been so used.

The failure to rely upon discourse as a key indicator of social

knowledge can be explained. Intelligent discourse cannot be conducted without specific knowledge of the topic at hand. The storehouse of isolated bits of knowledge possessed by a student is seen as an indicator of the student's potential for conducting intelligent discourse. Multiple choice and fill-in-the blank items, therefore, have been considered reasonable stimuli for generating indicators of a student's familiarity with the raw material that certain forms of discourse require. There are, however, several problems in relying upon students' choices among predetermined alternatives or their production of short answers to stem questions.

First, there is no absolute sense in which knowledge can be tested. Whether we have a defined unit of knowledge in memory and/or can retrieve it can be determined only in relation to the task that we choose as an indicator. Each of the following, for example, might be an indicator of whether a student can identify US Presidents in the 20th century: (a) Which of the following Presidents served in the 20th century? .... (b) Name three US Presidents who served in the 20th century. Suppose a student passed (a), but failed (b) What would we conclude about her knowledge of 20th century US Presidents?

Second, we question the assumption that the acquisition, storage and retrieval of isolated knowledge bits must occur prior to application and synthesis of knowledge in problem-solving and production of discourse. The fact that intelligent discourse displays or reveals specific content and topical knowledge does not require that such knowledge be learned or tested first in the form of isolated bits, separate from the discourse intended to illuminate a problem or topic. The belief that we first engage in the simple process of acquiring knowledge bits and later learn the more complex process of using them has been increasingly challenged by cognitive science. Instead, acquisition, storage, retrieval, application, synthesis are all seen as complex, interactive processes (Resnick, 1986; Romberg, 1986).

Third, is the familiar curricular and pedagogical danger that in the effort to adequately prepare students for discourse, we spend so much time transmitting knowledge that students rarely have an opportunity in school to produce discourse. Research indicates that retention of isolated knowledge bits is short lived unless the bits are frequently put to use. In this sense conventional testing provides unreliable indicators that can be expected to show vast declines over short periods of time, except for those particular items that students continue to use. This all-too-common result is often attributed to lack of teaching time, but its persistence leads one to suspect a more fundamental cause: educators and the public conceiving of knowledge itself primarily as retrieval of isolated knowledge bits rather than as the conduct of intelligent discourse.

Testmakers might accept these criticisms of the tendency to teach and to test for isolated knowledge bits, and they might show much enthusiasm for discourse as the ultimate indicator of knowledge. Their reluctance to emphasize discourse assessment may be based essentially on other issues; namely, problems of validity, reliability, comparability and the costs of discourse assessments. These issues will be addressed later.

III What forms of discourse should be assessed and why?

Having developed a general argument for using discourse as an indicator and criticized major assumptions that have stood in the way of more attention to discourse assessment, we now consider the forms of discourse that might be assessed. Within the fields of history and social studies, what, more precisely, should the nation expect students to speak about intelligently?

A. Selection of Common Content

Knowledge in this curriculum field consists of seven academic disciplines (history, geography, economics, political science, psychology, sociology, anthropology), and several other emerging bodies of knowledge (e.g. area studies, ethnic studies, environmental studies, global education, law-related education). Within and between fields, scholars and practitioners disagree about the central topics (among a virtually infinite set) on which all students should be able to write or speak.

Cultural pluralism within the US poses a further challenge for the selection of common content. Social studies, in contrast to subjects such as math or science which also experience controversy over selection of the central topics to be taught, is particularly concerned with helping students to become productive citizens within the communities to which they belong. To the extent that different racial, ethnic, gender, and socioeconomic groups experience different contexts and problems in the exercise of citizenship, it becomes increasingly difficult to define a single or common set of topics likely to maximize constructive democratic citizenship for all students in the nation.

Lack of consensus about critical knowledge within the broad field combined with unique knowledge needs among different groups in the exercise of citizenship can accentuate the influence of local schools and communities in defining curriculum unique to their interests. To help clarify national goals in the field (which be must done if there is to be a national assessment) and to facilitate comparative analysis, we should, nevertheless, continue the attempt to identify a discrete set of topics on which we would want all students to conduct intelligent discourse.

Selection of common content is not the topic of this paper. Presumably Fraenkel has addressed this issue in his, and most of the work on this issue should be undertaken by a future committee. Nevertheless, we might suggest a few possible topics likely to be endorsed by a substantial portion of the professional and lay community: The Founding of the United States; the US Constitution and Bill of Rights; Slavery; Rights and Responsibilities of Democratic Citizenship. Other efforts to outline critical topics should also be consulted; for example, National Council for the Social Studies (1979; 1984), Boyer (1983), College Board (1983; 1986).[2]

B. Elective Content

As a matter of principle and as a matter of necessity, it is unwise to prescribe that all or most of the instruction in social studies should be based in a national common curriculum. A case can be made for a limited common curriculum focusing on topics yet to be identified, but a national assessment should also offer an opportunity for students to demonstrate competence beyond the critical areas. Presumably, students devote significant energy to the study of some topics emphasized in their schools not emphasized by schools as a whole. This body of "uncommon" knowledge either required by schools in different ways or studied by students as electives will presumably influence the texture and direction of citizen discourse in the nation. We should gain some indication the extent to which students can speak and write intelligently about these topics as well. In addition to assessment of common topics, therefore, we should assess discourse also on a set of elective topics.

Further discussion of the selection of critical topics and the extent of elective topics may well lead to specification of a set of areas across which some portion of the elective topics would be distributed. It may be decided, for example, that each student should be able to conduct intelligent discourse on a topic in the history of the United States, on a topic in the history of another nation or culture, and on an elective topic from a broader set of possibilities.

C. Degree of Specificity in Content

---

[2]This discussion of what kind of discourse to assess assumes that a major feature of NAEP will be retained; namely, that items will be distributed among samples of students such that no individual student will be asked to show competence on all of the topics to be assessed. National indicators of the level of discourse can be made with different students addressing different topics.

The task of selecting general topics or elective topics should not include specification of discrete bits of knowledge, particular concepts, theories or explanations characteristic of conventional assessments. Since the goal here is intelligent discourse (not the recall or application of atomized knowledge), and since discourse on these topics can reasonably touch upon a variety of issues and rely upon a variety of specific knowledge, the assessment must provide considerable latitude for students in the kinds of specific knowledge they demonstrate in their statements. This commitment to topics, rather than to specified knowledge, is necessary not only to remain faithful to discourse as the ultimate objective, but also to recognize (even celebrate) the fact that different schools may teach different knowledge as they study the critical topics.

This is not to minimize the importance of substantive content in the evaluation of discourse. Intelligent discourse in history and social studies demands that students address and incorporate authoritative knowledge in their effort to construct meaningful statements about social life. Thus, discourse in social studies entails not simply the use of general language or thinking skills, but the production of "domain-specific" knowledge in the students' language. To evaluate student success in social studies discourse, it will be necessary, therefore, to develop general criteria for the content or knowledge that ought to emerge in the discussion of specific topics. Content guidelines, however, must be flexible and allow for variation among students.

D. Forms of Discourse: Written and Oral

Students should be able to produce intelligent discourse on social topics both orally and in writing. Each form is necessary for effective citizenship in a democracy. In either case, students can be presented with a topic, a general question, a problem, a provocative quotation, vignette or excerpt from a published work and asked to make a reasonably extended statement (e.g. about 300 words) in response. The response should indicate that the student can use knowledge to develop an explanation, an idea, a proposal, an argument, a narrative; and that various parts of the statement are organized into a coherent whole. The precise nature of these tasks and more specific criteria for assessing success on these general criteria must be developed. The Appendix contains a few illustrative tasks.

Since previous national assessments have not focused on writing unique to social studies, substantial development of format and evaluative criteria may be necessary. Extensive experience in the assessment of writing will be helpful, but new work on the evaluation of knowledge within discourse will be necessary.

Assessment of oral discourse should also be pursued. Oral dialogue is the form in which we most frequently demonstrate social knowledge to others. More importantly, it is the dominant way in which we express the concerns of citizenship. At first glance, it seems to raise two major problems. (a) Teachers rarely emphasize the production of intelligent oral discourse as the central goal of instruction. Thus, students are unaccustomed to developing extended oral statements as indicators of what they have learned. (b) Assessing oral discourse would seem to present new logistical problems in the recording and assessment of statements. Audiotape recording followed by manual transcription, or perhaps, in the future, by computerized voice-to-print transcription could be used. The more difficult problem, also relevant to written discourse, may be promoting this form of competence as a central objective in social studies classrooms.

IV What does experience and research suggest regarding the feasibility of this approach to assessment?

Recent reviews of testing practices advocate the need for departures from traditional multiple choice formats in social studies and other fields as well. Even in the fields of mathematics and science, experts have called for techniques that allow students more opportunity to explain their approaches to the solution of problems, rather than simply choosing from among predefined solutions.[3] In the United Kingdom, innovative assessment efforts in mathematics, science, foreign language, and "oracy" have all stimulated greater opportunities for students to converse, to produce language, and to complete more holistic tasks than conventional tests have permitted (Burstall, 1986).

In the NAEP 1981-82 assessment of citizenship and social studies, twenty five of the sixty four released exercises were open-ended, requiring students to write answers. Most of these exercises, however, called for short, one or two sentence answers, or the making of lists; for example "Give two things you could do to find out why each candidate thinks as he does." "Name as many reasons as you can why laws are needed." A few items asked students to explain a course of action and provided enough space to imply that more extended discourse was called for. High reliability in scoring such items was obtained. The criteria, however, emphasized primarily whether correct information was included in the answer and included little attention to the form or sophistication of the argument. For

---

[3]Alternative approaches to assessment in mathematics and science are suggested by Collis (in press), Romberg (1986), Fredericksen (1986) and the Committee on Indicators of Precollege Science and Mathematics Education (1986).

the most part, the exercises were tightly focused on particular units of knowledge, offering little opportunity for students to define a problem and to organize information from several sources to bear on it.[4] In developing more elaborate procedures for assessing discourse, however, we should build upon NAEP's experience with previous open-ended items.

The assessment of writing probably offers the most extensive experience on which to develop further efforts in social studies.[5] Current scholarship in the field of English (Greenberg, Wiener, & Donovan, 1986; Tyler, 1983; White, 1986) supports the efficacy, reliability, and validity of the collection and scoring of writing samples as a means of assessing students' skills in composition. Evidence of support for the use of writing samples for multiple purposes, including placement in English composition classes, determination of proficiency or minimal competency, and for diagnosis comes from three sources: the membership of the National Testing Network in Writing, the National Assessment of Educational Progress, and the National Writing Project.

The National Testing Network in Writing is an information clearinghouse formed in 1981 with funds from the Office of Academic Affairs of the City University of New York and the Fund for the Improvement of Post secondary Education. The National Testing Network in Writing conducted a 1983 survey of its membership concerning their attitudes and practices regarding writing assessment. The results of this survey (Greenberg, Wiener, & Donovan, 1986) indicate network membership support for the position that writing should be directly assessed via writing samples scored by trained faculty readers. Further, network members preferred locally developed, holistically scored writing samples over any other sort of test, regardless of test purpose.
The National Assessment of Educational Progress has conducted three national assessments of writing in the school years ending in 1974, 1979, and 1984. These assessments elicited writing samples from 95,000 students aged nine, thirteen, and seventeen. Three types of writing tasks, informative, persuasive, and imaginative, were constructed to reflect the varied purposes for which students write, at home, at school, and in the community. Both holistic scoring and primary trait scoring methods were used to assess students' writing strengths and to compare their performance across agemates and over

---

[4]Carlson (1986) estimated that only 8% of the 1981-82 NAEP objectives for social studies dealt with discourse, and his analysis of released items agrees with these points.

[5]The following material on writing assessment was developed by Mary Louise Gomez.

time (Applebee, Langer, & Mullis, 1986).

The National Writing Project, developed under the direction of Professor James Gray of the University of California-Berkeley, annually trains 70,000 teachers at more than two hundred sites located in the United States and abroad. The NWP is considered a significant effort at curriculum reform in the teaching of skills of written expression and in 1984 received an award for its exemplary university/school collaboration component. (This was jointly presented by the American Association for Higher Education and the Carnegie Foundation for the Advancement of Teaching.)

The National Writing Project advocates and teaches, in its summer institutes and teacher in-service programs, multiple models for evaluating student writing. These models include those of holistic and primary trait scoring. Miles Myers, for many years the Assistant Director of the Bay Area Writing Project, homebase for the National Writing Project, is also the author of the text Procedures for Writing Assessment and Holistic Scoring (published by the National Council of Teachers of English in 1980).

Despite such endorsements, some debate remains concerning the use of direct measures of writing assessment as opposed to multiple-choice assessment. Edward M. White, Professor of English at California State University-San Bernadino and director of the 1980-85 National Institute of Education--funded project concerning Research in Effective Teaching of Writing, summarizes the current status of the multiple choice versus essay testing of writing debate.

> The issue of multiple-choice versus essay testing of writing has become a bit dated since the development of relatively reliable and cost-effective holistic scoring procedures. Insofar as a test seeks to measure actual writing ability, the issue can be more generally defined as a conflict between direct and indirect measures (White, 1986, p. 61).

White and others (Fader, 1986; Greenberg, Wiener, & Donovan, 1986) argue that multiple-choice tests do measure some of the skills assessed through the collection and scoring of writing samples, including students' abilities to recognize conventions of grammar, sentence structure and mechanics, and to correctly choose from several stylistic options. However, multiple-choice tests do not directly measure those skills consistently identified by writing teachers as embodying composition. These skills include the invention, revision, and editing of ideas to serve particular

audiences and needs with appropriate stylistic, mechanical, and grammatical structures.

Until recently, the advocates of multiple-choice, indirect measures of assessing writing cited the high cost and low reliability of scoring writing samples. This argument has shifted due to the high development costs of multiple-choice testing, its needs for constant revision and security monitoring, questions concerning the validity of indirect measures, and the potential negative effect upon curricula due to lack of attention to actual writing (White, 1986).

Holistic scoring and primary trait scoring predominate as forms of direct assessment of students' writing skills. Holistic or general impression scoring is a guided procedure focusing on a writer's general fluency in responding to a writing task as well as attending to organization, quality of content, spelling, punctuation, and vocabulary choice. Primary trait scoring focuses on the specific features necessary to accomplish a particular writing task as well as the author's understanding of audience (Applebee, Langer, & Mullis, 1986). Both types of scoring are widely accepted as producing reliable scores at acceptable levels of cost.

Holistic evaluation is usually conducted by teams of trained readers using scoring guides describing desired features of the writing and identifying high, medium, and low quality levels for each feature. Primary trait evaluation is done in a similar manner with teams of trained readers using scoring guides which define and give examples of the particular features sought in each writing task. For example, the 1984 NAEP assessment elicited persuasive writing samples from students aged nine, thirteen, and seventeen. Students ages thirteen and seventeen were asked to write a letter to their principals requesting either the morning or afternoon session of a split school day. The following describes a minimal response to the task (as defined by the scoring guide).

> In order to achieve at least a minimal level, the writers needed to clearly take a stand and support it with at least one appropriate reason that was consistent with their point of view. The result could be quite short, as in the following example (Applebee, Langer, & Mullis, p. 22).

Each scoring guide contains descriptions and examples of papers at each of several levels of quality.

Estimates concerning the number of papers possibly scored by individual readers vary from two minutes per paper (Cooper, 1977) to twenty or more papers per hour for complex essays written in 45-minute increments (White, 1986) to 20-minute essays read at a rate

of thirty-eight per hour (Conlan, 1986). Cooper (1977) and Tyler (1983) argue that scoring reliability as high as .90 for individual writers can be achieved with this method. Standards of reliability are maintained through the development of good scoring guides accompanied by illustrative papers at each level as well as the development of a sense of community and a consensus to use the guides among readers (White, 1986).

The greatest cost of such assessment is the fee paid to readers. Members of the Wisconsin Writing Project regularly conduct such scoring for a fee of $50.00 for a four hour, half-day period per teacher. Some local school districts provide summer work in scoring for interested faculty members while other schools assign such tasks to an administrative team headed by a reading or language arts specialist.

In summary, direct assessment of students' writing skills is a widely accepted and regularly practiced process yielding data for multiple purposes, including student placement, determination of proficiency or minimal competency, and diagnosis.

One of the major challenges of discourse assessment in social studies will be to devise scoring schemes that tap not simply the skills of written and oral communication, but the students' incorporation of social studies knowledge into their language. The final indicator should be not merely a proxy for writing skill nor for correct knowledge, but a measure of success in integrating knowledge into one's use of language.

V Will Discourse Assessment Provide Meaningful Comparative Indicators?

Let us assume that feasibility problems in the administration and scoring of discourse can be solved such that for each task it will be possible to assign a proficiency score that represents the student's success in integrating significant social studies knowledge into a coherent statement. To the extent that the same tasks are given to students in different regions or states, or to same age students in different years, proficiency rates can be meaningfully compared.[6] Some might object that proficiency scores are so vague that they fail to allow comparison of precisely which bits of knowledge students include in their discourse. There are two answers to this objection. First, if the nation is interested in testing more precise units of knowledge, this can be accomplished

---

[6]Care must be taken, of course, to avoid unfair comparisions. For example, comparing discourse indicators of schools whose student bodies differ greatly in socio-economic background can give misleading conclusions regarding school quality.

through traditional measures; discourse assessment should not be asked to bear that burden. Second, some of the discourse tasks may legitimately require students to demonstrate relatively specific knowledge, but this could be cued in the directions for the task and included in scoring criteria. For example, illustration #2 in the Appendix might stipulate that students demonstrate knowledge of three branches of government, of at least two provisions of the Bill of Rights, and of election procedures for Congress. Such stipulations would offer more precise bases for comparison, but they must be approached cautiously to avoid transforming discourse tasks back into conventional tasks of recall and recognition.

Discourse analysis poses more of a challenge for meaningful comparisons across age groups. The social studies field is notorious for the problems it faces in prescribing appropriate knowledge for students at different ages. To define precise forms of discourse that should be demonstrated by students of different ages is even more complex. Illustrations in the Appendix were written with high school seniors in mind. For those illustrations, we started at the year of high school graduation and asked, "What forms of discourse on social topics should be demonstrated by young adults about to assume full responsibilities of citizenship?" It might be possible to identify an earlier transition point, perhaps 8th grade, and to identify more elementary discourse tasks on which the final tasks might build. Of course, there is no reason to require cross age comparisons on every indicator of social studies knowledge. Discourse on social topics might be one of those tasks assessed only toward the end of secondary school.

International comparisons will also be difficult, but not primarily because of the discourse format. What makes these comparisons problematic is the fact that nations will differ in the forms of history and social science knowledge they intend for their youth. Perhaps a survey of nations would reveal some commonalities, but I suspect these will be rare. Each nation attempts to teach students something of its own cultural heritage. Thus, specific content tested might be different in each nation. It is conceivable, however, that each nation might produce an index of youth understanding of its own history. These indicators could be compared, but interpretation would be difficult. Unless nations can agree on the need for discourse assessment and some common criteria for success for discourse in social studies, there will be no basis of comparison. Hopefully, the argument presented in II above establishes the case for discourse assessment as a national priority for the US, regardless of its potential for international comparison.

References

Adler, M.J. (1982). The paideia proposal: An educational manifesto. New York: Macmillan.

Applebee, A.N. (1984). Writing and reasoning, Review of Educational Research, 54:4, 577-596.

Applebee, A. N., Langer, J. A., & Mullis, I, V. S. (1986). NAEP: Writing trends across the decade, 1974-84. Washington, DC: National Assessment of Educational Progress.

Boyer, E.L. (1983). High school: A report on secondary education in America. New York: Harper & Row.

Burstall, C. (1986). Innovative forms of assessment: A United Kingdom perspective, Educational measurement: Issues and practice, 5:1, 17-22.

Carlson, K. (draft August 1986). The National Assessment of Educational Progress in Social Studies. New Brunswick, NJ: Rutgers University.

College Entrance Examination Board (1986). Academic preparation in social studies: Teaching for transition from high school to college. New York: CEEB.

College Entrance Examination Board (1983). Academic preparation for college: What students need to know and be able to do. New York: CEEB.

Collis, K.F. (in press). Levels of reasoning and the assessment of mathematical performance. In T.A. Romberg (Ed), The monitoring of school mathematics: Background. Madison, WI: Wisconsin Center for Education Research.

Committee on Indicators of Precollege Science and Mathematics Education, Commission on Behavioral and Social Sciences and Education, National Research Council (draft January 1986). Review of science content in selected student achievement tests.

Conlan, G. (1986). "Objective" measures of writing ability. In K. L. Greenberg, H. S. Wiener, & R. A. Donovan (Eds.), Writing assessment: Issues and strategies (pp. 109-125).

Cooper, C. R. (1977). Holistic evaluation of writing. In C.R.Cooper, & L. Odell, Evaluating writing: Describing, and measuring, judging (pp. 3-32). Urbana, IL: National Council of Teachers of English.

Fader, D. (1986). Writing samples and virtues. In K. L. Greenberg, H. S. Wiener, & R. A. Donovan (Eds.), Writing assessment: Issues and strategies (pp. 79-92).

Fredericksen, N. (1984). The real test bias: Influences of testing on teaching and learning, American Psychologist, 39:3, 193-202.

Fredericksen, N. (draft July,1986). VII. Indicators of student learning in mathematics and science. Princeton, NJ: Educational Testing Service.

Greenberg, K. L., Wiener, H. S. & Donovan, R. A. (Eds.). (1986). Writing assessment: Issues and strategies. New York: Longman, Inc.

Halliday, M.A.K., & Hasan, R. (1976). Cohesion in English. London: Longman.

McPeck, J.E. (1981). Critical thinking and education. Oxford: Martin Robertson.

National Council for the Social Studies Task Force on Scope and Sequence (1984). In search of a scope and sequence for social studies, Social Education, 48:4, 249-262.

National Council for the Social Studies. (1979). Social Studies curriculum guidelines. Washington, DC: NCSS.

Nystrand, M. (in press). The structure of written communication: Studies in reciprocity between writers and readers. New York: Academic Press.

Nystrand, M., & Wiederspiel, M. (1977). Case study of a personal journal: notes toward an epistemology of writing, in M. Nystrand (Ed) Language as a way of knowing. Toronto: Ontario Institute for Studies in Education.

Oliver, D.W,. & Shaver, J.P. (1966). Teaching public issues in the high school. Logan, UT: Utah State University.

Resnick, L.B. (draft April 1986). Education and learning to think. Report for the Commission on Behavioral and Social Sciences and Education, National Research Council.

Romberg, T.A. (draft June,1986). Measures of mathematical achievement: problems and influences. Madison, WI: University of Wisconsin.

Sizer, T.R. (1984). Horace's compromise: The dilemma of the American high school. Boston, MA: Houghton Mifflin.

Tyler, R. W. (1983). Testing writing: Procedures vary with purposes. In R. W. Bailey, & R. M. Fosheim, Literacy for life: The demand for reading and writing (pp. 197-206). New York: Modern Language Association.

White, E. M. (1986). Pitfalls in the testing of writing. In K. L. Greenberg, H. S. Wiener, & R. A. Donovan (Eds.), Writing assessment: Issues and strategies (pp. 53-78). New York: Longman Inc.

Appendix

Illustrative Tasks for Assessing Student Discourse

1. The 18th century is considered one of the most important periods in the history of the United States. Explain some ways why that century is considered so important. In expressing your ideas, describe key events, people, general trends, ideas that seemed to take hold of people, and explain why these would be considered important.

2. A person from outer space (who speaks English) arrives on the planet and asks, "What kind of government does the United States have?" How would you answer the visitor who has never been exposed to democratic institutions? Include an explanation of how laws are passed, enforced and judged, and what rights and responsibilities the citizens have.

3. Americans have been accused of being ignorant of other parts of the world and isolated from other cultures. Show that you are innocent of this charge by describing a society, culture or country outside of the United States. Discuss as much as possible of its history, geography, economy, culture, and critical issues it faces today.

4. What do you think of the following statement? "Americans owe most of their main beliefs, values and institutions to the Western European tradition." Explain how much you agree or disagree with this statement and provide evidence for your position.

5. Joe is 25 years old and out of work. He graduated from high school, and has worked most of his life in construction, but he lives in a region where construction workers have been laid off by the thousands. He asks why there aren't more jobs. He is told only that "We're in a recession," or "Interest rates are too high," or "Demand is depressed." Joe never studied economics, and finds these statements hard to understand. Explain to Joe, in language he is likely to understand, several possible reasons why the economy may not be providing more jobs for construction workers.

6. A friend is coming to town by car to visit you. He phones from the highway on the outskirts of the city, asking for directions on how to find your home. Use the map below to give verbal directions on the shortest way to travel from his arrival point to your home. [The map would include a variety of streets and landmarks, with several possible routes.]

7. The table below gives information on population and age distribution trends in different parts of the US, economic production, unemployment, employment in different vocations, and

educational attainment. [insert an appropriate table] Using this information, describe major changes that occurred in the United States between 1900 and 1960 and also describe the variables, if any, that have remained relatively stable for long periods of time. Wherever possible explain the reasons for change and stability.